What is BIDS?

• It’s a data **structure**; nothing to do with **format** per se
• It’s about: - how you organize data in a folder
  - how you name files
  - how you document metadata
  - using community standards and dictionaries to do all
    the above
• It cares about imaging data but also behaviour/cognition

bids.neuroimaging.io
BIDS is a widely supported “best practice” (becoming \textit{a posteriori}, the “standard”, rather than been created \textit{de novo})
BIDS is a widely supported “best practice”

https://github.com/bids-standard
The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments


Data Sharing Task Force

The Poldrack Lab @ Stanford
Evolution of BIDS

1. Kickoff meeting at Stanford in Spring 2015
2. Meeting at OHBM 2015 (June)
3. Introduced to neuroinformatics community at INCF Congress 2015 (August)
4. First release candidate and public call for comments (September)
5. Version 1.0.0 published along the introductory paper

→ Initially covered structural MRI and fMRI, now all sort of MRI, PET, EEG, MEG, iEEG, extensions for animal, connectivity, imaging genomics ..
Datatype specific publications

EEG


iEEG


MEG

The BIDS goal is to make more data accessible to more researchers
Making more data accessible

- for yourself in 6 months time

- to people in the lab
  (new students, collaborators, governance)

- to other researchers
  (data sharing)
Meet Prof. Smith

I need to ask Mike about these ‘old’ data, there is so much more we can do with this.
Meet Mike

A Data Scientist is...

A Business Analyst that lives in California.
Prof Smith is lost in her data
Getting lost in your data

- **Multitude of techniques** (MEG, EEG, PET, MRI, NIRS, TMS, etc ...) and applications.

- Despite similarities in experimental designs and data types **each researcher tends to organize and describe their data in their own way.**
BIDS principles
Principles

1. Adoption is crucial
   = got to fit peoples' needs = driven by user cases

2. Don't reinvent the wheel
   = use already in use file formats = minimize work for users

3. 80/20 rule
   = 80% of the work is already done by you collecting data, just need 20% to get it to BIDS
BIDS data structure basics
A BIDS folder (any/all modality)

- source (optional)
- stimuli (optional)
- derivatives (optional)
- sub-XXX

Anyone can now find his/her way around data!
Metadata as text files (tsv, json) with standard dictionary

<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>stimuli</td>
<td>21/03/2018 21:58</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-002</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-003</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-004</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-005</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-006</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-007</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-008</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-009</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-010</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-011</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-012</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-013</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-014</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-015</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-016</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-017</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-018</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-019</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>dataset_description.json</td>
<td>15/03/2018 11:30</td>
<td>JSON File</td>
</tr>
<tr>
<td>participants.tsv</td>
<td>19/03/2018 20:21</td>
<td>TSV File</td>
</tr>
<tr>
<td>README.txt</td>
<td>15/03/2018 11:33</td>
<td>TXT File</td>
</tr>
</tbody>
</table>

```
{
    "Name": "",
    "BIDSVersion": "",
    "License": "",
    "Authors": "",
    "Acknowledgements": "",
    "HowToAcknowledge": "",
    "Funding": "",
    "ReferencesAndLinks": "",
    "SourceDatasetsURLs": ""
}
```

Your data are identifiable and citable
Metadata as text files (tsv, json) with standard dictionary

<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>stimuli</td>
<td>21/03/2018 21:58</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-002</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-003</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-004</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-005</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-006</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-007</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-008</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-009</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-010</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-011</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-012</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-013</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-014</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-015</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-016</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-017</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-018</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>sub-019</td>
<td>24/03/2018 08:56</td>
<td>File folder</td>
</tr>
<tr>
<td>dataset_description.json</td>
<td>15/03/2018 11:30</td>
<td>JSON File</td>
</tr>
<tr>
<td>participants.tsv</td>
<td>19/03/2018 20:21</td>
<td>TSV File</td>
</tr>
<tr>
<td>README.txt</td>
<td>15/03/2018 11:33</td>
<td>TXT File</td>
</tr>
</tbody>
</table>

participant_id  age  sex
sub-002      34   M
sub-003      12   F
sub-004      33   F

Subjects info are shared at the root – easy to figure populations, age, and other basic demographics

Human and machine readable!
A BIDS folder (any/all modality)

Name redundancy – harder to make mistakes; files come with metadata (json)

Inside folder sub-002
  → anat folder
    sub-002_T1w.nii.gz
  → eeg folder
    ????
A BIDS folder (any/all modality)

Name redundancy – harder to make mistakes; files comes with metadata (json)

Inside folder sub-002
  → anat folder
    sub-002_T1w.nii.gz

  → eeg folder
    sub-002_something_eeg.set
BIDS for EEG

EEG-BIDS, an extension to the brain imaging data structure for electroencephalography

Cyril R. Pernet, Stefan Appelhoff, Krzysztof J. Gorgolewski, Guillaume Flandin, Christophe Phillips, Arnaud Delorme & Robert Oostenveld

The Brain Imaging Data Structure (BIDS) project is a rapidly evolving effort in the human brain imaging research community to create standards allowing researchers to readily organize and share study data within and between laboratories. Here we present an extension to BIDS for electroencephalography (EEG) data, EEG-BIDS, along with tools and references to a series of public EEG datasets organized using this new standard.
EEG File formats (known and supported)

Anyone can read the data
- official: edf & vhdr/eeg
- unofficial: set/fdt, bdf
  (all supported by open-source software)
(i) EEG File formats

(i) wide usage in the community

(ii) open access documentation, open source implementation for both reading and writing in at least two programming languages and widely supported in multiple software packages (both open source and commercial)

(iii) high numerical precision (16 and 32 bits respectively).

Some (M/iE/E)EG specifics metadata
Some (M/iE/E)EG specifics metadata
Some (M/iE/E)EG specifics metadata

Yes, this info is often present on a channel
- as triggers, no metadata/name
- also easier to figure out as user
  (i.e. yourself in 6 months) what those event codes are
Prof. Smith (2030)
What other benefits?
Benefits

• Use each other data / well documented / ready to analyse
• You can have **pipelines!**

![Diagram showing workflow and pipeline steps]

- **Workflow**: the sequence of computational steps through which a piece of work passes from initiation to completion.
- **Pipeline**: set of data processing elements connected in series, where the output of one element is the input of the next one – pipelines implement workflows.
From BIDS-formatted EEG data to sensor-space group results: a fully reproducible workflow with EEGLAB and LIMO EEG
EEGLAB-LIMO pipeline

Fully reproducible code from import BIDS to figures

Attend our session this Thursday to check this out
How do I start?
Preparing data using BIDS
Preparing data using BIDS
Preparing data using BIDS
Preparing data using BIDS
Sharing data using BIDS

Findable
Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Accessible
Data is deposited in a trusted repository.

Interoperable
(Meta)data uses a formal, shared, and broadly applicable language or format.

Reusable
Data is described with clear and understandable attributes. There should be a clear and acceptable license for re-use.

https://www.force11.org/group/fairgroup/fairprinciples